To appear in the *Journal of Statistical Computation and Simulation*
Vol. 00, No. 00, Month 20XX, 1–17

Bayesian P-splines and advanced computing in R for a changepoint analysis on spatio-temporal point processes (Special Issue GRASPA)

L. Altieri[a*], D. Cocchi[a], F. Greco[a], J.B. Illian[b,c] and E.M. Scott[d]

[a]*Department of Statistical Sciences, University of Bologna, Italy;* [b]*School of Mathematics and Statistics, University of St Andrews, CREEM, UK;* [c]*NTNU, Trondheim, Norway;* [d]*School of Mathematics and Statistics, University of Glasgow*

(*Received 00 Month 20XX; final version received 00 Month 20XX*)

This work presents advanced computational aspects of a new method for changepoint detection on spatio-temporal point process data. We summarise the methodology, based on building a Bayesian hierarchical model for the data and declaring prior conjectures on the number and positions of the changepoints, and show how to take decisions regarding the acceptance of potential changepoints. The focus of this work is about choosing an approach that detects the correct changepoint and delivers smooth reliable estimates in a feasible computational time; we propose Bayesian P-splines as a suitable tool for managing spatial variation, both under a computational and a model fitting performance perspective. The main computational challenges are outlined and a solution involving parallel computing in R is proposed and tested on a simulation study. An application is also presented on a dataset of seismic events in Italy over the last 20 years.

Keywords: earthquake data; changepoint analysis; spatio-temporal point processes; spatial effect; log-Gaussian Cox processes; Bayesian P-splines; parallel computing

AMS Subject Classification: 62H11; 62M30

1. Introduction

This work deals with changepoint analysis on spatio-temporal point patterns. In this context, the traditional temporal changepoint analysis needs to be extended to spatio-temporal datasets where dependence over time, as well as over space, is allowed. A first approach to this problem has been proposed in [1]. This work is a fundamental step further in changepoint analysis as it allows changepoint analysis methods to be applied to complex spatio-temporal datasets where the assumption of independence between observations within the same time segments is unrealistic. Very little work has previously been done in this direction, only concerning temporal point processes [2].
In this paper, some further computational issues are addressed, that mainly concern modelling strategies for the spatial effect and the ability to obtain results in a reasonable time. A case study is also presented as an application example, which consists of the whole set of the main seismic events recorded over the last 20 years in Italy. An analysis of the Italian seismicity by means of a space-time branching model suited for

*Corresponding author. Email: linda.altieri@unibo.it

spatio-temporal point processes has been proposed by [3] and has also been applied by [4].

In spatial point process analysis, the aim is often to model the intensity surface as a function of spatial covariates and additional spatial random effects [5]. When a spatial point pattern is observed at different points in time, the evolution of the process and potential changes in its parameters are also of interest. Different classes of point process models have been proposed in order to take into account general features of the data generating process. Inhomogeneous Poisson processes are suitable when no interaction between points is assumed and when the intensity function, i.e. the limit of the expected number of occurrences when the considered area tends to zero [6], is non-stochastic, though allowed to vary over space. Gibbs processes model patterns where the direct interaction between points is the main determinant of the spatial distribution. Cox processes generalize Poisson processes as they assume a spatially structured driver that determines the point distribution over space, and that the intensity function is random; conditional on the intensity value at a specific location, the process follows a inhomogeneous Poisson process. In most cases, the intensity surface is assumed to be smooth, thus when a spatial effect is included it can be conveniently modelled as a Gaussian Field (GF) [7].

In this work, we consider Log-Gaussian Cox Processes (LGCPs) [6], an extremely flexible class of point process models, which have proven to be very effective in the context of ecological studies [8]. They are a subclass of Cox point processes, a broad set of spatial (and spatio-temporal, see Section 3) point process models that can be defined in two steps:

$$\begin{aligned} \Lambda &= \int_W \lambda_u du \\ X|\Lambda &\sim Poi(\Lambda) \end{aligned}$$

where $W \subset \mathbb{R}^2$ is the observation window and λ_u is the first order intensity function of the process X at a specific location $u \in W$. The intensity $\{\lambda_u\}_{u \in W}$ is a random field which can only take non-negative values, then its log-transformation is modelled as GF. Despite their flexibility and their suitability for many real situations, LGCPs have not been much used until very recent years. The problem is that, except for very special cases, the density of X is analytically intractable [9], and has to be approximated; indeed, the general form of the Cox process likelihood involves integration over the distribution of Λ which has infinite dimension [10]. The traditional approach for estimating a LGCP [11] consists in approximating its likelihood with a Poisson likelihood, by superimposing a grid over the window and counting the number of points Y_s in each cell C_s, with $s = 1, \ldots, S$. This way, the following Cox process is obtained

$$\begin{aligned} \Lambda_s &= \int_{C_s} \lambda_u du \\ Y_s|\Lambda_s &\sim Poisson(\Lambda_s) \end{aligned}$$

where usually the integral in Λ_s is impossible to compute and an approximation is needed: approximately, $Y_s \sim Poisson(|C_s|\lambda_s)$, where $\log(\lambda_s)$ is a representative value of the (continuous) GF within the cell C_s.

Commonly, the recommendation is to make the grid "sufficiently fine". This is usually not precisely defined, and leads to some interesting questions, concerning both computational time and smoothness of the estimates.

As regards computational time, finer grid resolutions correspond to increasing computational complexity. In a general model where both fixed effects (such as covariates) and random effects are included, the computational burden mainly concerns the spatial effect. The estimation of a spatial effect induces a computational complexity depending on the dimension of the grid.
As for the smoothness of the estimates, it is to consider that many situation require a certain degree of smoothing since rough estimates are too detailed to capture the spatial structure of the phenomenon under study.
The paper is organised as follows: Section 2 gives an outline on Bayesian changepoint analysis with focus on the challenges presented by an extension to spatio-temporal data; a new methodology for detecting changes is presented. Section 3 introduces some models for spatio-temporal point processes, highlights issues concerning the estimation of the spatial effect and proposes a solution based on Bayesian P-splines. Section 4 presents some advanced techniques for speeding up computations in R. Section 5 shows a simulation study to assess the validity of the proposed approach. Then, Section 6 provides an application to earthquake data. Finally, some concluding remarks are given in Section 7.

2. Bayesian changepoint analysis for spatio-temporal point processes

A changepoint is defined as a point θ in a data series Y such that the observations follow one distribution, say F_m, up to that point and another one, say F_{m+1}, after that point [12]. The assumptions are that data are ordered from 1 to T (usually in time order, but some other natural orders might be considered) and that the generating process presents some abrupt changes; data are then split into segments, which generally follow the same model but under different parameter specifications [2]. A further common assumption in standard changepoint analysis is that observations are i.i.d. within every time segment, therefore the distribution of the sequence can be written as $Y_t \sim F_j$, for $\theta_j < t \leq \theta_{j+1}$, where $j = 0, \ldots, M$ counts the number of changepoints, and $\theta_1, \theta_2, \ldots, \theta_M$ are the changepoint locations, defined in our work as the last time point of every segment (with the convention $\theta_0 = 0$ and $\theta_{M+1} = T$). When spatio-temporal changepoint analysis is considered, two main issues arise. Firstly, at every time point the datum consists of a realisation of a spatial point process, therefore different types of change over time may occur, since a change over time in the pattern can involve the average intensity of the process (expected number of points), the spatial distribution or both. A method is needed that is able to detect any of these changes. Secondly, in many cases it is not reasonable to assume independence between observations; this implies a non-tractable segment marginal likelihood and the need for approximate techniques. The work presented in [1] aims at answering these questions with a new Bayesian approach for spatio-temporal point process data with a inhomogeneous intensity function and dependence within time segments. The method has been assessed in a thorough simulation study, and has been shown to be able to detect different types of change. The use of Integrated Nested Laplace Approximation (INLA) [13] to estimate spatio-temporal models and to compute the segment marginal likelihoods makes the approach computationally tractable.
The method consists in choosing a model and fitting it repeatedly to the dataset assuming different changepoint positions. Let $\boldsymbol{Y}$ be the $(T \times S)$-dimensional data vector, where Y_{ts} denotes the number of points observed at cell s and time t, and $\boldsymbol{Y}_t$ denotes data observed on the grid at time t; hence, $\boldsymbol{Y} = (\boldsymbol{Y}_1', \ldots, \boldsymbol{Y}_t', \ldots, \boldsymbol{Y}_T')'$. Every time a changepoint is assumed at a time point $\theta = 1, \ldots, T$, the data vector is split into two segments $\boldsymbol{Y}_{t \leq \theta}$ and $\boldsymbol{Y}_{t > \theta}$, where $\boldsymbol{Y}_{t \leq \theta} = (\boldsymbol{Y}_1', \ldots, \boldsymbol{Y}_\theta')'$ and $\boldsymbol{Y}_{t > \theta} = (\boldsymbol{Y}_{\theta+1}', \ldots, \boldsymbol{Y}_T')'$. The model is

fitted separately to the two segments (independence across, though not within, segments is assumed). Two segment marginal log-likelihood values are obtained and summed to give the data marginal log-likelihood conditional on θ, $l(\boldsymbol{Y}|\theta) = l(\boldsymbol{Y}_{t\leq\theta}|\theta) + l(\boldsymbol{Y}_{t>\theta}|\theta)$. This is computed for $\theta = 1, \ldots, T$ to obtain a vector of marginal log-likelihoods

$$\boldsymbol{l} = (l(\boldsymbol{Y}|\theta = 1), \ldots, l(\boldsymbol{Y}|\theta = T))'.$$

This includes the case where $\theta = T$, i.e. no changepoint is present and data are thought as coming from the same generating process for $t = 1, \ldots, T$. The posterior distribution of the changepoint location is obtained via Bayes' Rule by multiplying the log-likelihood vector for a vector of prior probabilities over the changepoint positions. Once a posterior probability is obtained for every time point, decisions must be made as to which changepoints are to be accepted. To this aim, assuming a uniform prior on the changepoint location θ, we follow [1] and propose to choose a changepoint θ^* as the value of θ corresponding to the greatest marginal log-likelihood value; in particular, if $\theta^* = T$ no changepoint is detected in the dataset. Discussion about this criterion can be found in Section 7.

For a multiple changepoint search, a binary segmentation algorithm can be implemented as in [12], i.e. an iterative procedure which looks for a single changepoint for the whole dataset and, if found, iteratively splits the data at the changepoint dealing with the resulting segments separately until no more changes are detected in any segment.

A general Bayesian changepoint model needs prior specification on number and positions of the changes and a hierarchical model for the data segments. In the most general case, we look for an unknown number of changes at unknown time points. We take a uniform prior for the number $m = 0, \ldots, M$ of changepoints and we assume a minimum segment length of d time points when we wish to avoid unrealistic adjacent changes. Considering that changepoints are looked for sequentially, our prior can be written as

$$\begin{aligned} \pi(m) &= \frac{1}{M+1} \text{ for } m = 0, \ldots, M \\ \pi(\theta_1, \ldots, \theta_m|m) &= \pi(\theta_m|\theta_{m-1}, m)\pi(\theta_{m-1}|\theta_{m-2}, m) \ldots \pi(\theta_1|m) \end{aligned} \tag{1}$$

where $\pi(\theta_1|m) = (T - 2 \times d)^{-1}$.

The conditional priors for $\theta_2, \ldots, \theta_m$ can be computed sequentially as the binary segmentation algorithm proceeds.

3. Modelling spatio-temporal point processes

In this Section, we illustrate our approach to spatio-temporal modelling of point processes. Much of the theory of spatio-temporal point processes comes from that of spatial point processes. General methods for the analysis of spatio-temporal point processes are not well established yet [14]. A spatio-temporal point process is defined in a subset $W_{tu} \subset \mathbb{R} \times \mathbb{R}^2$ where t is a time index, $t \in T \subset \mathbb{R}$, which can be continuous or discrete and $\boldsymbol{u} = (u_1, u_2)$ is a space index, $\boldsymbol{u} \in U \subset \mathbb{R}^2$.

A spatio-temporal LGCP can be defined as a spatio-temporal inhomogeneous Poisson process conditional on a stochastic intensity function λ_{tu} that varies both in space and time, with $\log(\lambda_{tu})$ following a Gaussian process. The spatio-temporal LGCP is extremely flexible as it enables the presence of both fixed and random effects [15].

With the grid approximation, the count at cell s and time t is modelled as:

$$Y_{ts}|\lambda_{ts} \sim Poisson(\lambda_{ts}|C_s|),\ t = 1, \ldots, T,\ s = 1, \ldots, S.$$

The choice of the grid implies a trade-off between computational complexity and accuracy of the approximation. We adopt a regular grid composed of squared cells, i.e. $|C_s| = |C|\ \forall s$. We manage spatial dependence by means of Intrinsic Gaussian Markov Random Fields (IGMRF). These models are defined by their precision matrix, that is sparse since it represents conditional dependence which, as reflected in the Markov structure, is based on a small number of neighbours. Moreover, IGMRFs are widely adopted for modelling lattice data, which is the type of data we generate by superimposing a grid over the window and taking the cell counts as response: in fact, each λ_{ts} is a representative value of the intensity over a grid cell rather than an instantaneous value. This approach has been much used over recent years (see, e.g., [8, 16]) as it allows reasonably complex models such as LGCPs to be fitted to point pattern data. It is now a standard approach for INLA users, but is also possible with traditional MCMC methodology.
The log-intensity is modelled as follows:

$$\log(\lambda_{ts}) = \beta_0 + \phi_t + \psi_s \tag{2}$$

where β_0 denotes the intercept term, while ϕ_t and ψ_s denote temporal and spatial random effects respectively. The inclusion of linear effects to capture dependence of the intensity on covariates is straightforward, but is beyond the aim of our work, which focuses on detecting changes in the intensity over space and/or time. In what follows $\text{IGMRF}_H(\tau \boldsymbol{K})$ denotes the multivariate normal distribution of a H-dimensional random vector with $(H \times H)$-dimensional structure matrix $\boldsymbol{K}$, a sparse matrix whose non-zero pattern describes conditional dependencies in the field, and precision parameter τ. Temporal and spatial random effects are modelled as:

$$\begin{aligned} \boldsymbol{\phi}|\tau_\phi &\sim IGMRF_T(\tau_\phi \boldsymbol{K}_\phi) \\ \boldsymbol{\psi}|\tau_\psi &\sim IGMRF_S(\tau_\psi \boldsymbol{K}_\psi). \end{aligned}$$

Model hierarchy is completed by prior specification for the hyperparameters. Following [17], we scale $\boldsymbol{K}_\phi$ and $\boldsymbol{K}_\psi$ in order to have the same marginal variance, and the same $Gamma(a, b)$ hyperprior is specified for the precision parameters τ_ϕ and τ_ψ. When modelling spatio-temporal point patterns, very often $T << S$, since a high resolution grid is needed for accurate likelihood approximation; thus, the computational complexity of the model is strictly related to S, i.e. to the dimension of the spatial effect. In Section 3.1, we propose an approach based on Bayesian P-splines that is motivated by the following rationale: we keep a fine grid resolution in order to obtain an accurate likelihood approximation and reduce the dimension of such spatial effect by using a set of spline bases on a knot grid with dimension $P < S$. This way, the computational burden is substantially reduced.
The advantage of this approach is twofold: on the one hand it allows computations to be speeded up, on the other hand it delivers a straightforward way to manage the random effect's smoothness. As a matter of fact, the problem with specifying an IGMRF on the whole grid is that it generally produces a rather detailed surface that may result in worse model performances with respect to smoother surfaces.

3.1. *Managing the spatial effect via Bayesian P-splines*

Splines are a powerful non-parametric approach to fitting curves to sets of data. The P-spline method has been first proposed by [18] as an efficient and computationally stable approach for smoothing, which can be extended quite naturally to multi-dimensional smoothing as shown in [19]. A general P-spline model consists of regression on a basis of B-splines with a regularization penalty on the spline coefficients. B-splines are required to be equispaced over the covariate domain, i.e. defined in correspondence of knots laying on a regular grid, in such a way that smoothing is regulated by the type of penalty adopted [20]. Typical penalties consider first or second order differences between neighbouring spline coefficients. The extension of such methods in a Bayesian framework has been proposed by [21], using a random walk prior on the spline coefficients to impose smoothness and a Gamma prior on the precision of the random walk. P-splines have been used as a general tool to model smooth surfaces in several contexts involving spatial data [22–26] but to the best of our knowledge they have not been used for modelling point process data so far.
By means of Bayesian P-splines, the spatial effect in model (2) is modelled as:

$$\psi_s = \boldsymbol{B}(s)^{'}\boldsymbol{\gamma} \tag{3}$$

where $\boldsymbol{B}(s)^{'}$ is the s-th row of a $(S\times P)$-dimensional basis matrix and $\boldsymbol{\gamma}$ is a P-dimensional vector of spline coefficients. A practical way to obtain the full basis matrix $\boldsymbol{B}$ is to use tensor products of marginal bases [27] defined over longitude and latitude coordinates (x_{1s}, x_{2s}) of the grid cells centroids, $s = 1, ..., S$. Marginal bases matrices $\boldsymbol{B}_j$, $j = 1, 2$, of dimension $S \times P_j$, are defined as a collection of P_j univariate B-splines centred at equally-spaced knots over each direction. In order to define a regular knot-grid over direction j, the interval $\boldsymbol{r}_j = [min(\boldsymbol{x}_j), max(\boldsymbol{x}_j)]$ can be divided into P_j' intervals of the same length by using $P_j' + 1$ equally spaced knots. A univariate B-spline is made of $g + 1$ polynomial pieces connected at the knots, each piece of degree g. We set $g = 3$, i.e. we adopt cubic splines. B-splines are numerically stable because they are local, i.e. non-zero over a limited domain spanned by $g+1$ intervals. The number of columns of the marginal basis matrix $\boldsymbol{B}_j$ is $P_j = P_j' + g$. For details on computing B-splines see [18]. Since the centroids of the approximation grid form a regular grid, the full basis matrix $\boldsymbol{B}$ can be easily obtained by Kronecker product of the marginal bases, i.e. $\boldsymbol{B} = \boldsymbol{B}_1 \otimes \boldsymbol{B}_2$, such that $P = P_1 \times P_2$. An example of how to superimpose knots on the approximation grid is shown in Figure 1, left panel, while the right panel displays a 3D plot of the resulting spline basis.
An IGMRF prior on spline coefficients $\boldsymbol{\gamma}$ is specified as:

$$\begin{aligned} \boldsymbol{\gamma}|\tau_\gamma &\sim \text{IGMRF}_P(\tau_\gamma \boldsymbol{K}_\gamma) \\ \tau_\gamma &\sim Gamma(a, b) \end{aligned}$$

where the structure matrix $\boldsymbol{K}_\gamma$ is a sparse precision matrix, whose non-zero entries reflect the conditional dependencies among spline coefficients. This prior was proposed by [21] for a univariate Bayesian P-spline model. The degree of variability of the joint posterior distribution of $\boldsymbol{\gamma}$ depends on both the structure matrix $\boldsymbol{K}_\gamma$ and the precision parameter τ_γ. According to the Markov assumption, the coefficients in $\boldsymbol{\gamma}$ are conditionally independent given coefficients at neighbouring knots. The definition of neighbourhood determines different smoothing models; see [28] for a detailed description of IGMRFs on regular lattices. We build the structure matrix $\boldsymbol{K}_\gamma$ by the Kronecker sum of marginal

structures $\boldsymbol{R}_j$, $j = 1, 2$ [19]. Let $\boldsymbol{R}_j$ be the structure matrix associated to a P_j-dimensional IGMRF defined on the knots in the j-th direction,

$$\boldsymbol{K}_\gamma = \boldsymbol{R}_1 \oplus \boldsymbol{R}_2 = (\boldsymbol{I}_{P_2} \otimes \boldsymbol{R}_1) + (\boldsymbol{R}_2 \otimes \boldsymbol{I}_{P_1}) \tag{4}$$

where $\boldsymbol{I}_q$ is the identity matrix of dimension q. A different degree of smoothness along each direction can be imposed by either assuming different $\boldsymbol{R}_j$'s or two precision parameters such that the precision matrix of the IGMRF is specified as $\tau_1 \boldsymbol{R}_1 \oplus \tau_2 \boldsymbol{R}_2$.

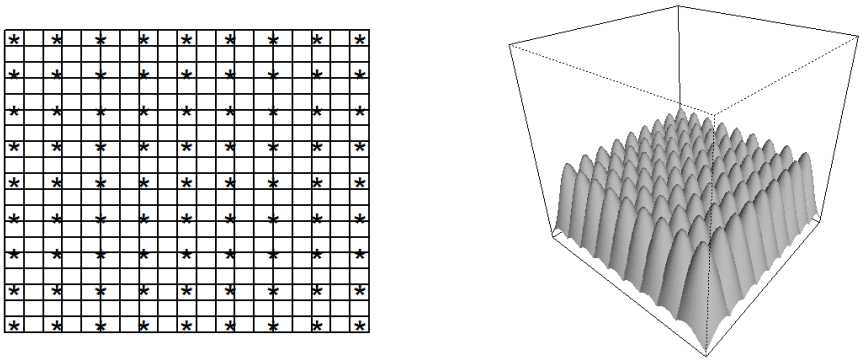

Figure 1. Approximation grid and knots to build the spline basis (left panel); bivariate spline basis (right panel).

4. Computations

INLA [13] is an alternative approach to MCMC for fitting Bayesian hierarchical models; it is not based on sampling (which is the key to its speed), it may be used to fit latent Gaussian models and the estimates turn out to be very exact in practice for a wide range of commonly used latent Gaussian models [13]. Moreover, tools for assessing the approximation error are provided. The approach produces precise estimates in seconds and minutes, even for models involving thousands of variables, in situations where any MCMC computation typically takes hours or even days [13]. This also means different models can be run and compared within reasonable time.
The reference software for working with INLA is `R` [29, 30]: the `R-INLA` package is used for all the computations in this work.
Despite the speed of the INLA approach and the further computational saving of using splines, it can still take a prohibitive time to obtain results of the changepoint analysis presented in Section 2 and 3. Indeed, we are dealing here with spatio-temporal point processes, where typically thousands of grid cells are used to approximate the likelihood at each time point, and the time series can potentially be very long. Therefore, we make use of some advanced computing techniques.
In this work, we make substantial use of the `R` package `parallel` [31] which contains much of the work done for packages `snow` and `multicore` [32] and is meant for parallel computing on physical CPU (Central Processing Unit). The CPU cores typically share some information: they share access to the total RAM and may share parts of the cache memory. Multicore CPUs are what the package `parallel` is designed to exploit [31].
Some basic parallel computing is already implemented in `R`. Indeed, a CPU core can have

more than one thread to speed up computations. A thread of execution is the smallest sequence of programmed instructions that can be managed independently [33]. Multiple threads can exist within the same process, executing concurrently and share resources such as memory. On a multicore system, multiple threads can be executed in parallel, with each of the processors or cores executing a separate thread simultaneously; on a processor or core with hardware threads, separate software threads can also be executed concurrently by separate hardware threads. Several `R` packages, such as `R-INLA`, use multiple threads.

Advanced parallel computing consists in running much larger pieces of computations at the same time. This approach is way faster than a `for` loop, where each single iteration only starts when the previous one has been completed. The key point to parallel computing is that the pieces of computations are unrelated (which is the case of some, but not all, loops); examples are evaluating the same `R` function or fitting the same model on different datasets, which can be particularly useful in simulation studies [34].

The basic approach sets up M different so-called 'worker' processes and splits the main task in M subtasks, which are sent to the workers by a 'master' process. When the workers are all done with computing, they are asked for results by the master. Often, the number of tasks is greater than the number of workers that can be created. In this case, the first M tasks are sent to the workers, then the next M are sent to the workers once they all complete their first task, and so on. A more sophisticated approach is called 'load balancing' and dynamically allocates the remaining tasks to the workers that complete the previous tasks first. This approach is preferred when the tasks are known to require different amounts of computational time, to avoid waiting; however, it is not advisable to have many more tasks than workers (such as by a factor of 10 or more).

The pool of workers is called a cluster of nodes; a number of approaches can create the cluster, here we present the two most common techniques:

- the first one is implemented via the function `makePSOCKcluster()`[31]. It creates a number of working processes that communicate with the master via sockets. A socket is an endpoint of an inter-process communication across a computer network, defined by an address, the combination of an IP address and a port number (much like one end of a telephone connection is the combination of a phone number and a particular extension). Based on this address, internet sockets deliver incoming data packets to the appropriate application process or thread. Note that the workers need to be completely initialized as they do not share the master's workspace.
- The second way is implemented via the function `makeForkCluster()` (not available for Windows) [31]. This creates nodes which are exact copies of the master process, including the workspace, and is very fast.

The number of nodes for the parallel work must be decided based on several some factors such as:

- the number of cores available. The function `detectCores()` tries to determine the number of cores available in the machine;
- the number of tasks to run for each node, which should not be too great (as said for the load balancing approach);
- the type of task that needs to be run. For example, when working with processes that use multiple threads themselves (such as `inla` calls) a small number of nodes such as 2 or 4 are to be preferred in order to enjoy the maximum computational advantage.

The machine we use for our computations has a RAM of 128 GBs, 16 cores and 32 threads. Thus, when a core uses 100% of its memory it uses 8 GBs, but it can use up to

1600% of its memory, i.e. the total RAM. We set a cluster of 4 nodes via forking.
Once the size of the cluster of nodes is set, the main functions of `parallel` to run are parallelized and more sophisticated versions of `apply`, `lapply` and `sapply` functions [31].

5. Simulation study

A preliminary simulation study was designed to assess the improvement in computational time of parallel algorithms over sequential algorithms, using model (2) on a short time series and a small grid. We verified that time drops substantially when we switch from a sequential algorithm like a `for` loop to fit the models under different changepoints to parallel computing with an increasing number of nodes. As explained in Section 4, we benefit from a greater improvement in time if we stick to a small number of nodes. In particular, our first simulation study needs an average computational time of 47 seconds for a sequential algorithm, 22 seconds when using a cluster of 2 nodes, 14 seconds for a cluster of 4 nodes and 10 seconds for a cluster of 8 nodes.
In what follows, we dicuss a simulation study meant to compare the changepoint detection ability and the accuracy of models (2) and (3); all computations have been run on a cluster of 4 nodes.
We generate $G = 100$ realizations from the following model:

$$Y_{tsg} \sim Poisson(\lambda_{tsg})$$
$$\log(\lambda_{tsg}) = \beta_0 + \widetilde{\phi}_{tg} + \widetilde{\psi}_s$$

$t = 1, \ldots, T$, $s = 1, \ldots, S$, $g = 1, \ldots, G$, where we set $T = 15$ and $S = 2500$, i.e. we use a grid of 50×50 cells; the changepoint is located at time $\widetilde{\theta} = 7$. The intercept β_0 is set to $\log(1)$ for $t \leq \widetilde{\theta}$ and to $\log(3)$ for $t > \widetilde{\theta}$. This implies that the expected number of points in a single cell is 1 before $\widetilde{\theta}$ and 3 afterwards. The temporal effects $\widetilde{\boldsymbol{\phi}}_{t\leq\widetilde{\theta},g}$ and $\widetilde{\boldsymbol{\phi}}_{t>\widetilde{\theta},g}$ are generated from an AR(1) model with autoregressive parameter $\rho = 0.95$. The spatial effect is specified as a smooth surface over the grid cell centroid standardized coordinates as

$$\widetilde{\psi}_s = \alpha_1 \sin(2\pi x_{1s}) + \alpha_2 \sin(2\pi x_{2s})$$

where $\alpha_1 = \alpha_2 = 0.35$. The generated surface is displayed in Figure 2, left panel. The right panel shows the number of points over the pattern, $\sum_s Y_{tsg}$, for each time point t and realization g; the change in the intensity of each pattern before and after $\widetilde{\theta}$ is due to the shift in β_0. We looked for a changepoint in the series using the IGMRF approach as in model (2) and the P-spline approach as in model (3) with 15×15, 25×25 and 35×35 knots.

5.1. *Simulation results*

In this Section, where not specified, results and comments refer to the 15×15-dimensional knots grid; only some differences are reported for other knot grid resolutions.
As far as the computational time is concerned, as expected the P-spline approach has been faster; the total time for 100 simulations has been close to 48 hours for the IGMRF approach and 25 hours for the P-spline approach with 15×15-dimensional knots.

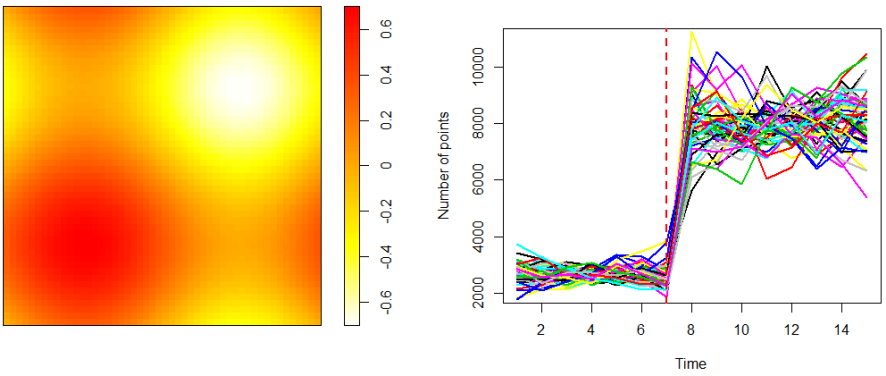

Figure 2. Spatial effect (left panel); number of points for each time series, 100 simulations (right panel).

Table 1. Changepoint detection performance over 100 simulations

	IGMRF	**P-Splines 15x15**	P-Splines 25x25	P-Splines 35x35
correct changepoint detection ($\theta = 7$)	86%	**92%**	88%	86%
Δ marginal log-likelihood	-4196	—	-717	-1815
	(-4200;-4191)	—	(-719;-715)	(-1818;-1812)

Table 2. Estimated intercept (exp scale).

	IGMRF		P-Splines 15x15		P-Splines 25x25		P-Splines 35x35	
	Seg 1	Seg 2	Seg 1	Seg 2	Seg 1	Seg 2	Seg 1	Seg 2
mean	1.000	3.000	1.001	3.000	1.001	2.979	1.001	2.979
2.5 %	0.987	2.973	0.985	2.960	0.986	2.970	0.987	2.972
97.5%	1.011	3.019	1.024	3.044	1.018	3.031	1.015	3.025

The ability to detect the correct changepoint location is shown in the first line of Table 1. As can be seen, the P-spline approach with a 15×15-dimensional knots grid shows the highest percentage (92%) of correct detections while other models deliver a slightly worse performance. As for comparison of model fitting performances, the average marginal log-likelihood for each model is obtained as:

$$\bar{l} = G^{-1} \sum_{g=1}^{G} l_g(\boldsymbol{Y}|\theta_g^*)$$

where θ_g^* denotes the changepoint location detected for the g-th replicate. The highest marginal log-likelihood value is obtained under the P-spline model with a 15×15-dimensional knots grid. This value is chosen as a reference in the second line of Table 1, where values for competing models are reported as differences with respect to the reference value. It can be seen that increasing the dimension of the random field capturing spatial variation causes increasingly worse fitting performances.

Posterior estimates for the intercept over the simulated dataset are summarised in Table 2 on the natural scale: as can be seen, they are very accurate with respect to the original values 1 and 3. The empirical confidence interval becomes slightly larger as the spline basis dimension decreases.

As for the temporal effect, hardly any difference can be found between the IGMRF

and the P-spline approach. Estimates for the spatial effect are also very similar among approaches, but a slight increase can be observed in the estimated surface roughness when increasing the dimension of the random fields capturing spatial variation. Indeed, the variance of the spatial effect obtained with the P-splines is smaller than the variance resulting from the IGMRF approach in 92% of the replicates.
Our simulation studies allow some conclusions to be drawn:

(1) both IGMRF and P-spline methods show a satisfactory ability to detect the correct changepoint location (see Table 1, line 1);
(2) the P-spline approach delivers better results in terms of model fitting: increasing the dimension of the random field produces a decrease in the marginal log-likelihood (Table 1, line 2);
(3) all estimates (intercept, temporal and spatial effect) are extremely similar between the two approaches;
(4) adopting the P-spline approach allows for a substantial reduction of computational time, particularly when the number of knots is considerably smaller than the number of grid cells used for likelihood approximation;
(5) there is a substantial time saving in running parallel computations with a small cluster of nodes (typically 2 or 4 nodes).

Therefore, the P-spline approach has two main advantages in this context: faster computations and better fitting performance. Further comments can be found in Section 7.

6. Case study

In this Section we apply changepoint analysis to a collection of seismic events over the Italian territory. Data are made available by the National Institute of Geophysics and Vulcanology (INGV) and can be downloaded at `http://terremoti.ingv.it/it/`. Data come from 390 monitoring stations located over Italy, which operate 24 hours a day, 7 days a week, and are published in real time; for each event, the spatial coordinates, the hypocentre depth and the magnitude are reported. We analyse a set of 13254 events of magnitude 2.5 and above (earthquakes below this limit are not felt by people). The study period ranges from January, 1995 to December, 2014. An overall map of the hypocentre locations is presented in Figure 3. We split the dataset into yearly patterns and obtain a time series of $T = 20$ spatial point patterns with a number of seismic events ranging from 304 to 1592, with an average of 663 per year.

A changepoint analysis can answer many questions concerning the evolution of the seismic phenomenon over the Italian territory. Specific issues that need to be addressed are listed in many recent articles on the INGV website (`http://www.ingv.it/`) and highlight concerns about changes occurring in the distribution and magnitude of earthquakes (see the work in [35, 36]). A changepoint analysis is a suitable method for the analysis of the seismic events and since it is reasonable to assume spatial correlation and temporal dependence among the events, the development of a methodology that can model such a complex dataset will help us to assess if changes have occurred in the period between 1995 and 2014.

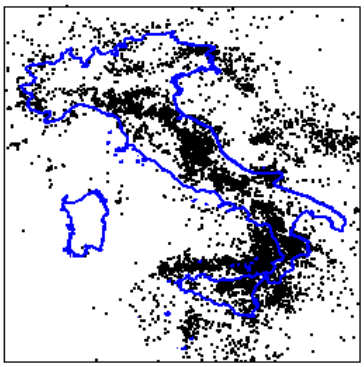

Figure 3. Seismic events of magnitude≥2.5, 1995-2014.

6.1. *Implementation and results*

Considering that $T = 20$, we look for one main changepoint θ, i.e. $M = 1$ and $m = 0, 1$; we take a minimum segment length $d = 3$. Following the prior setting in (1), we specify a uniform prior on both presence and location of the changepoint: $\pi(m) = 2^{-1}$ and $\pi(\theta|m = 1) = 14^{-1}$. The same models (2) and (3) as in the simulation study are fitted repeatedly to the dataset in order to find the most likely changepoint location. The approximation grid is made of 80×80 cells ($S = 6400$): this is the highest feasible resolution for computations under the IGMRF approach, since higher resolutions caused a crash in the R-INLA program. With the P-Splines approach, we keep the same resolution for comparability reasons, but we successfully tested the approach with higher resolution grids. The number of knots for the cubic splines is 25 in each cardinal direction, delivering a set of $P = 625$ spline coefficients to be estimated; $P << S$ i.e. the number of knots used for estimating the random effect with the P-spline approach is much smaller than the number of grid cells used for both likelihood approximation and estimation of the spatial random effect following the IGMRF approach. A different number ($P = 1600$) was also tested, but a 25×25 dimensional knots grid is preferable in this context in terms of marginal likelihood values and computational time.
As in Section 5, all model fitting is done using R-INLA and parallel computing on a cluster of 4 nodes.
The detected changepoint location is $\theta^* = 2008$ with both the IGMRF and the P-spline approach. It is to remember that the changepoint is defined as the last point of a time segment (Section 2), therefore $\theta^* = 2008$ means the change in the parameters took place in 2009. This result is supported by knowledge about recent seismic events in Italy (see Section 7 for further comments).
As in the simulation study, the marginal log-likelihood values are markedly higher when using the P-spline approach. This holds both for the whole vector of log-likelihood values as in Equation 2 and specifically for $l(\boldsymbol{Y}|\theta = \theta^*)$; therefore, it is again suggested that the P-spline approach, and consequently the corresponding estimate for the spatial effect, should be favoured.
In the application under study, the estimation of the spatial effect is fundamental, as the main interest lies in understanding if there is any change in the spatial behaviour of earthquakes. Indeed, once the approximation grid cell area is included in the model as an offset, hardly any difference can be detected for both approaches in the intercept and temporal effect estimates before and after θ^*. Moreover, the smoothness of the spa-

tial random effect estimated surface is the other main reason for preferring P-splines to the IGMRF approach. The estimate obtained using P-splines (reported in Figure 4) is smooth and it is possible to appreciate the change after 2008, which consists in the presence of two main hotspots in the centre of Italy, corresponding to Emilia-Romagna and Abruzzo areas.

The third advantage of P-splines in this context (beyond log-likelihood values and

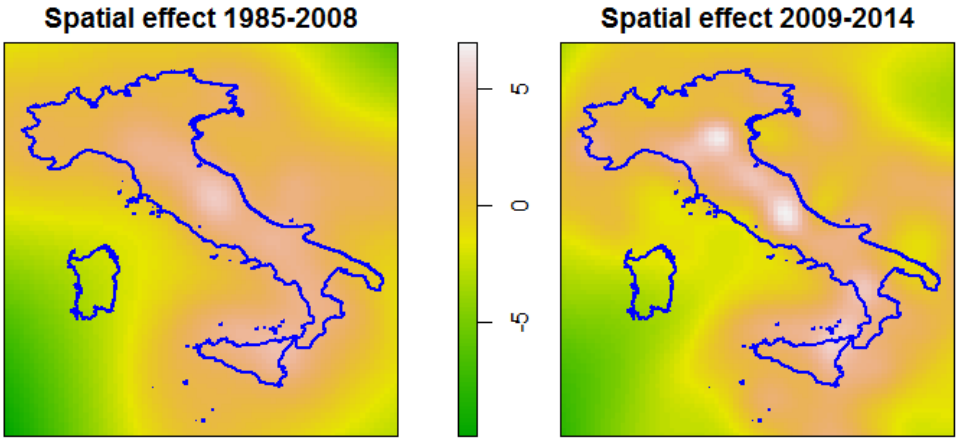

Figure 4. Estimated spatial effect before and after the changepoint using P-splines.

smoothness of the estimates) is the computational time saving. Indeed, the time needed for the analysis is less than half of the time used by the IGMRF approach. Intuitively, the time saving decreases as P increases, nevertheless the use of a rich basis such as a 40×40 dimensional knots grid ($P = 1600$) still brings a reduction of 20% of the time.

7. Concluding remarks

This work is aimed at solving some advanced computational issues concerning the detection of changepoints over spatio-temporal point pattern series, and the ability to produce good estimates for the spatial random effect. We reviewed the new methodology proposed in [1], based on a Bayesian approach to the changepoint detection and on the exploitation of INLA to obtain marginal log-likelihood values for the data segments. Then, we introduced two main tools: Bayesian P-splines, a flexible and powerful approach for the estimation of random effects, and parallel computing, a set of advanced techniques meant to speed up the production of results. An advantage of the work presented here is that it can be extended beyond point processes to a general analysis of any spatio-temporal dataset where a smooth latent process is assumed to drive the data. Indeed, the response data are areal counts.

This work allows some concluding remarks.

Our Bayesian methodology for changepoint detection has proven capable to detect temporal changes in spatio-temporal point processes. The inclusion of dependence within, though not across, segments, along with the methodology in [1], allows an extension of changepoint analysis methods to spatio-temporal data where an abrupt change in the model parameters takes place. When a change occurs, it is reasonable to assume that the data come from a different model. Indeed, most of the work on traditional temporal

changepoint analysis [12] relies on the assumption of independence between changes. An interesting further extension would be to include dependence across segments. We refer to [37] for pioneering this methodology for temporal series.

Moreover, we are able to investigate if the change concerns average intensity, spatial distribution or both. Indeed, with our method we can decompose the intensity estimation, which is often looked at as a whole, into fixed and random effects and study them separately. If our goal were solely to consider the number of events over the observation window per time point, then a traditional time series analysis would be appropriate. However, our objective is not simply the number of events but especially their spatial distribution. Our methodology has been developed in the most general case to consider multiple parameters to describe the point process.

A further point regarding our detection approach is our focus on the marginal log-likelihood. In Bayesian inference, the posterior ratio should be theoretically used for model choice; the problem is that the data marginal likelihood is needed to compute the Bayes Factor inside the ratio, which is usually very hard to obtain. Thus, surrogates such as the DIC are used instead. Nowadays, with INLA it is possible to obtain accurate approximation of the marginal log-likelihood. Therefore, in absence of prior knowledge about the changepoint number and location, marginal log-likelihood values can be used to decide if there is a changepoint and where it lies.

The main step beyond the methodology proposed in [1] is the inclusion of Bayesian P-splines, which brings three substantial advantages. Firstly, better model fitting with respect to the IGMRF approach in [1] based on the marginal log-likelihood values. As far as model performance is concerned, there is no substantial difference between the two approaches when it comes to detecting the correct changepoint locations, where both methods have proven effective. The key difference lies in the log-likelihood values, as in both the simulation and the application values produced by the P-spline approach are markedly higher than the ones delivered by the IGMRF approach. Secondly, P-splines are intrinsically able to deliver smooth estimates, which is fundamental in a context where the basic assumption is that there is a smooth field determining the point spatial distribution. When the underlying driver of the process is assumed to be smooth, detailed estimates are not always desirable, since they often happen to look 'noisy' and might miss the underlying spatial trend. Estimates obtained using an IGMRF model might therefore be considered too rough. On the other hand, the estimation of the fixed effects' coefficient is more accurate on a finer grid. The P-spline approach we propose allows fixed and random effects to be estimated separately. Splines allow to keep an extremely high grid resolution; this implies a small error with regard to the true point location, avoids ecological fallacy and produces very accurate estimates for all fixed effects. The desired smoothness for the random part can be obtained independently from the data grid, since the estimation of the spatial effect is based on a knots grid. Lastly, P-splines bring a considerable computational time reduction due to the smaller dimension of the random field. Even if the GMRF approach has a huge computational saving with respect to other methods, for large datasets computations can still be relatively slow and, in the end, even infeasible. When spatio-temporal data are taken into account, the computational time becomes a crucial issue for long time series, and the need to reduce it as much as possible is a main concern.

The need to save computational time also led us to the inclusion of advanced computational techniques globally labelled as parallel computing. We enjoyed a considerable further reduction in the computational time, and we believe this set of techniques can be of interest to any application where multiple independent tasks need to be run in order to obtain results.

As regards the application, results are consistent with some events of public knowledge: after 2008 two major seismic events in L'Aquila and in the Emilia-Romagna region shocked Italy. Indeed, there is a change in the spatial distribution: until 2008 earthquakes were evenly distributed all along the Appennini; afterwards, a clusterisation has taken place around the central-east part of Italy (where Emilia-Romagna and L'Aquila are) and, secondarily, around the volcanic islands close to Sicily. In the application context, the time unit is the year; this means that a changepoint in 2008 leads to a different behaviour in the earthquake intensity and pattern starting from 2009, which is indeed the year of the Abruzzo earthquake sequence. The year seems a reasonable unit in the context of earthquake data, where the underlying causes of seismic events typically span over a very long time range; for this reason, the choice of a smaller time unit could lead to the detection of small scale variation over time, which is not of interest here. Moreover, point pattern analysis is sensible when there is a sufficient number of points for every pattern, which is only the case when we group many observations together. However, any time scale may be chosen if preferred, and the changepoint can potentially be located more precisely over time, as the available dataset reports the exact date of the events.
As in several applications, it would be of interest to include extra knowledge (such as covariates or informative priors) in order to improve the reliability of the results. Useful information regards number and sensitivity of the monitoring stations and their evolution over time. The detection of earthquakes is related to the distance from the hypocentre and to the magnitude of the event; it might be of interest to investigate whether an increased density of the process might be partially due to an increased ability to record seismic events. Moreover, the depth of the hypocentre may be exploited in order to check if it is negatively correlated to the earthquake magnitude; besides, a changepoint analysis of the depth itself may bring useful knowledge to the interpretation of the phenomenon. This case study may benefit from a further extension of the methodology to dependence across temporal segments; at the moment, this is the most general available technique for changepoint analysis on spatio-temporal point processes. In addition, the focus of this paper is more on solving computational issues of the methodology itself, rather than on the case study.
Further methodological work might focus on tuning the hyperprior on the precision parameter of the random field in order to control the resulting level of smoothness.

Funding

As regards authors Linda Altieri and Fedele Greco, the research work underlying this paper was partially funded by a FIRB 2012 grant (project no. RBFR12URQJ; title: Statistical modeling of environmental phenomena: pollution, meteorology, health and their interactions) for research projects by the Italian Ministry of Education, Universities and Research.

References

[1] Altieri, L., Scott, M., Cocchi, D., Illian, J. (2015). A changepoint analysis on spatio-temporal point processes. *Spatial Statistics*, http://dx.doi.org/10.1016/j.spasta.2015.05.005

[2] Wyse, J., and Friel, N. and Rue, H. (2011). Approximate simulation-free Bayesian inference for multiple changepoint models with dependence within segments. *Bayesian Analysis*, **6**, 501-528

[3] Adelfio, G. and Chiodi, M. (2015). FLP estimation of semi-parametric models for space-time point processes and diagnostic tools. *Spatial Statistics*, doi:10.1016/j.spasta.2015.06.004.
[4] Nicolis, O. Adelfio, G. and Chiodi, M. (2015). Windowed ETAS models with application to the Chilean seismic catalogs. *Spatial Statistics*, doi:10.1016/j.spasta.2015.05.006.
[5] Møller, J., Waagepetersen, R. (2006). Modern statistics for spatial point processes. Aalborg: Department of Mathematical Sciences, Aalborg University. (Research Report Series; No. R-2006-12)
[6] Diggle, P. J. (2014). Statistical analysis of spatial and spatio-temporal point patterns. Third Edition. Taylor & Francis Group, Boca Raton
[7] Liang, S., Carlin, B.P., Gelfand, A.E. (2009). Analysis of Minnesota colon and rectum cancer point patterns with spatial and nonspatial covariate information. *The Annals of Applied Statistics*, **3**, 943-962
[8] Illian, J. B. (2012). A toolbox for fitting complex spatial point process models using Integrated Nested Laplace Approximation (INLA). *The Annals of Applied Statistics*, **6**, 1499-1530
[9] Waagepetersen, R. (2008). Estimating functions for inhomogeneous spatial point processes with incomplete covariate data. *Biometrika*, **95**, 351-363
[10] Diggle, P.J., and Moraga, P., and Rowlingson, B. and Taylor, B. M. (2013). Spatial and spatio-temporal log-Gaussian Cox Processes: extending the geostatistical paradigm. *Statistical Science*, **95**, 542-563
[11] Møller, J. and Waagepetersen, R. P. (2004). Statistical Inference and Simulation for Spatial Point Processes. Chapman and Hall/CRC, Boca Raton
[12] Chen, J. and Gupta, A. K. (2012). Parametric statistical change point analysis. Birkhauser, Boston
[13] Rue, H., Martino, S., Chopin, N. (2009). Approximate Bayesian inference for latent Gaussian models by using Integrated Nested Laplace Approximations. *Journal of The Royal Statistical Society B*, **71**, 319-392
[14] Cressie, N. and Wikle, C. K. (2011). Statistics for spatio-temporal data. John Wiley & Sons, Hoboken
[15] Taylor, B. M., and Davies, T. M., and Rowlingson, B. S. and Diggle, P. J. (2013). Lgcp: inference with spatial and spatio-temporal log-Gaussian Cox Processes in R. *Journal of Statistical Software*, **52**, Issue 4
[16] Yue, Y.R., Loh, J.M. (2011). Bayesian semiparametric intensity estimation for inhomogeneous spatial point processes. *Biometrics*, **67**, 937-946
[17] Sørbye, H. S., Rue, H. (2014). Scaling intrinsic Gaussian Markov random field priors in spatial modelling. *Spatial Statistics*, **8**, 39-51
[18] Eilers, P.H.C., Marx, B.D. (1996). Flexible smoothing with B-splines and penalties. *Statistical Science*, **11**, 89-102
[19] Eilers, P.H.C., Currie, I.D., Durbán, M. (2006). Fast and compact smoothing on large multidimensional grids. *Computational Statistics & Data Analysis*, **5**, 61-76
[20] Eilers, P.H.C., Marx, B.D. (2010). Splines, knots and penalties. *Wiley Interdisciplinary Reviews: Computational Statistics*, **2**, 637-653
[21] Lang, S., Brezger, A. (2004). Bayesian P-splines. *Journal of Computational and Graphical Statistics*, **13**, 183-212
[22] Goicoa, T., Ugarte, M.D., Etxeberria, J., Militino, A.F. (2012). Comparing CAR and P-spline models in spatial disease mapping. *Environmental Ecological Statistics*, **19**, 537-599
[23] Lee, D.J., Durbán, M. (2009). Smooth-Car mixed models for spatial count data. *Computational Statistics & Data Analysis*, **53**, 2968-2977
[24] Ugarte, M. D., Goicoa, T., Militino, A. F. (2004). Spatio-temporal modeling of mortality risks using penalized splines. *Environmetrics*, **21**, 270-289
[25] Bruno, F., Greco, F., Ventrucci, M. (2015). Non-parametric regression on compositional covariates using Bayesian P-splines. *Statistical Methods and Applications*, DOI 10.1007/s10260-015-0339-2
[26] Lee, D.J., Durbán, M. (2011). P-spline ANOVA-type interaction models for spatio-temporal

smoothing. *Statistical Modelling*, **11**, 49-69
[27] Wood, S.N. (2006). Generalized Additive Models: An Introduction with R. Chapman and Hall/CRC
[28] Rue, H., Held, L. (2005). Gaussian Markov Random Fields. Chapman and Hall/CRC
[29] Lindgren, F., Rue, H. (2015). Bayesian spatial modelling with R-INLA. *Journal of Statistical Software*, **63**, Issue 19
[30] Blangiardo, M., Cameletti, M. (2015). Spatial and spatio-temporal Bayesian models with R-INLA. Wiley, Chichester
[31] R-core Team (2010). Package 'parallel'. Available online at https://stat.ethz.ch/R-manual/R-devel/library/parallel/doc/parallel.pdf
[32] Urbanek, S. (2009-present). Multicore: parallel processing of R code on machines with multiple cores or CPUs. Available online at https://cran.r-project.org/package=multicore
[33] McCallum, Q.E., Weston S. (2011). Parallel R. Data analysis in the distributed world. O'Reilly Media
[34] Schmidberger, M., Morgan, M., Eddelbuettel, D., Yu, H., Tierney, L., Mansmann, U. (2009). State of the art in parallel computing with R. *Journal of Statistical Software*, **31**, Issue 1
[35] Chen, K.P., Tsai, Y.B., Amorese, D., Chang, W.Y. (2011). Incorporating change-point detection updates of frequency-magnitude distributions within the Taiwan Earthquake Catalog. *Terrestrial, Atmospheric and Oceanic Sciences*, **22**, 261-269
[36] Shearer,P.M., Stark, P.B. (2012). Global risk of big earthquakes has not recently increased. *Proceedings of the National Academy of Sciences*, **109**, 717-721
[37] Fearnhead, P., Ziu, L. (2011). Efficient Bayesian Analysis of Multiple Changepoint Models with Dependence across Segments. *Statistics and Computing*, **21**, 217-229